Exploratory Analysis of Rainfall Data in India for Agriculture Machine Learning

Submitted By

L.Hirithik kalin

B.Dineshmani

S.Mohammed Baseer

N.Sakthi Saravanan

TEAM ID - PNT2022TMID18172

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

MEPCO SCHLENK ENGINEERING COLLEGE, SIVAKASI

(An Autonomous Institution affiliated to Anna University Chennai)

November 2022

Institution Mentor: Dr.E.Amurtha

Industrial Mentor: Lalitha Gayathri

1. INTRODUCTION

1.1 Project Overview

Predicting weather events to stop loss of life to humans and the environment due to changing climatic conditions and irregular weather patterns is a critical issue facing humanity. Since there have been significant climatic changes in recent years, appropriate preventive actions are required. Floods may occur as a result of heavy rain. Flash floods can be disastrous. Disasters are occurring more frequently, more intensely, and with greater magnitude due to climate change. As a result, more According to the UN Office for Disaster Risk Reduction, weather-related phenomena like heatwaves, storms, floods, and droughts have been responsible for 90% of severe disasters during the past 20 years (UNISDR). Natural disasters are happening more frequently and with greater force. Predictions and emergency preparedness are challenging because of changing weather patterns. Thus, by conducting an exploratory study of the data gathered, we concentrate on the accurate estimation of the likelihood that a flood would occur in a certain place and the recommendation of a nearby evacuation area.

1.2 Purpose

To plan a disaster management system by predicting a flood event in order to reduce flood risk by suggesting an evacuation location from flood hazard zones, which eventually aids in managing the environment and water resource system. By building a model and choosing the most accurate prediction algorithm from the available classifiers, this also serves an early warning system function. The occurrence of flash floods has the potential to be extremely harmful to society. People who live close to riverbeds are first and foremost affected. providing them with the shelter they required and removing them from the danger zones. It has been challenging to anticipate the

occurrence of floods using conventional approaches, which has resulted in enormous destruction, due to the irregular change in climate patterns. Consequently, new approaches are developed to deal with severe situations and deal with flash floods.

In order to lessen the damage that a flash flood would cause, technology needs to be more thoughtful. It is now considerably simpler to anticipate floods and provide safe havens for evacuees in the current day. The rainfall readings are collected, combined from many resources, curated, mined, evaluated, and predictions are made over patterns in hazardous locations that are vulnerable to destruction and devastation. The prognosis is accompanied with a list of society-related recommendations. Early warning systems are included into climate change adaption strategies.

2. LITERATURE SURVEY

2.1 Existing Solutions

In this research, Input parameter like average temperature in month, wind velocity, humidity, and cloud cover was conceder for predicting rainfall in non-monsoon session. The performance of the results was measure with MSE (mean squared error), correlation coefficient, coefficient of efficiency and MAE (mean absolute error). The results of SVR were compared to those of MLP and simple regression technique. MLP being a computationally intensive method, SVR could be used as an efficient alternative for runoff and sediment yield prediction under comparable accuracy in predictions.[1]

In this research, Different ground-based weather features such as Temperature, Relative Humidity, Dew Point, Solar Radiation, PWV along with Seasonal and Diurnal variables are identified, and a detailed feature correlation study is presented. While all features play a significant role in rainfall classification, only a few of them, such as PWV, Solar Radiation, Seasonal, and Diurnal features, stand out for rainfall prediction. Based on these findings, an optimum set of features are used in a data-driven machine learning algorithm for rainfall prediction.[2]

The study contributes to using various classification algorithms for rainfall prediction in the different ecological zones of Ghana. The classification algorithms include Decision Tree (DT), Random Forest (RF), Multilayer Perceptron (MLP), Extreme Gradient Boosting (XGB) and K-Nearest Neighbour (KNN). The dataset, consisting of various climatic attributes, was sourced from the Ghana Meteorological Agency spanning 1980 – 2019. The performance of the classification algorithms was examined based on precision, recall, f1-score, accuracy and execution time with various training and testing data ratios.[3]

In general, weather and rainfall are highly non-linear and complex phenomena, which require progressive computer modeling and simulation for their precise prediction. Numerous and diverse machine learning models are used to predict the rainfall which are Multiple Linear Regression, Neural networks, K-means, Nave Bayes and more. These systems implement one of these applications by extracting, training and testing data sets and finding and predicting the rainfall.[4]

In this paper, we propose a new forecasting method that uses a deep convolutional neural network (CNN) to predict monthly rainfall for a selected location in eastern Australia. To our knowledge, this is the first time applying a deep CNN in predicting monthly rainfall. The proposed approach was compared against the Australian Community Climate and Earth-System Simulator-Seasonal Prediction System (ACCESS), which is a forecasting model released by the Bureau of Meteorology. In addition, the CNN was compared against a conventional multi-layered perceptron (MLP). The better mean absolute error, root mean square error (RMSE), Pearson correlation (r), and Nash Suttcliff coefficient of efficiency values were obtained with the proposed CNN.[5]

2.2 Reference

1. Xiaobo Zhang;Sachi Nandan Mohanty;Ajaya Kumar Parida;Subhendu Kumar Pani;Bin Dong;Xiaochun Cheng. **Annual and Non-Monsoon Rainfall Prediction Modelling Using SVR-MLP: An Empirical Study From Odisha**. (IEEE)(2020)

2. Shilpa Manandhar;Soumyabrata Dev;Yee Hui Lee;Yu Song Meng;Stefan Winkler; **A Data-Driven Approach for Accurate Rainfall Prediction**. (IEEE)(2019)

3. Nana Kofi Ahoi Appiah-Badu;Yaw Marfo Missah;Leonard K; Amekudzi;Najim Ussiph;Twum Frimpong;Emmanuel Ahene; **Rainfall Prediction Using Machine Learning Algorithms for the Various Ecological Zones of Ghana**.(IEEE)(2021)

4. Deepali Patil Shree L.R; **Rainfall Prediction using Linear approach & Neural Networks and Crop Recommendation based on DecisionTree**.(IEEE)(2020)

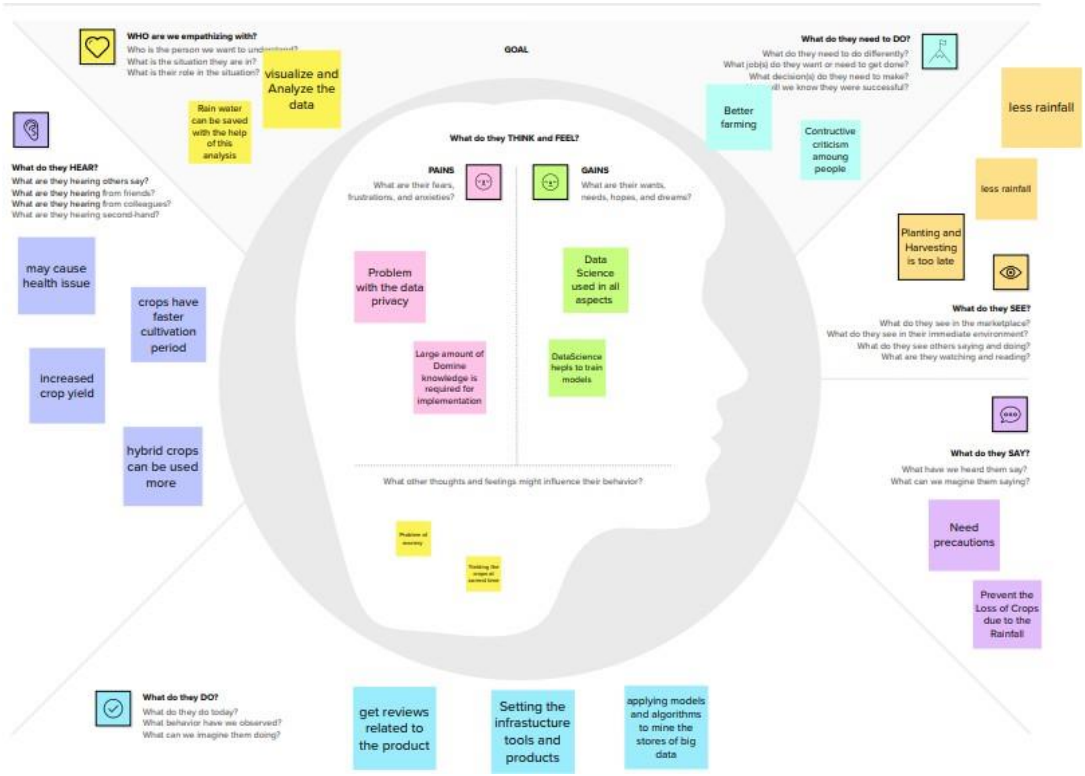
5. A. Haidar, B. Verma; **Monthly Rainfall Forecasting Using One-Dimensional Deep Convolutional Neural Network** .

2.3 Problem Statement Definition

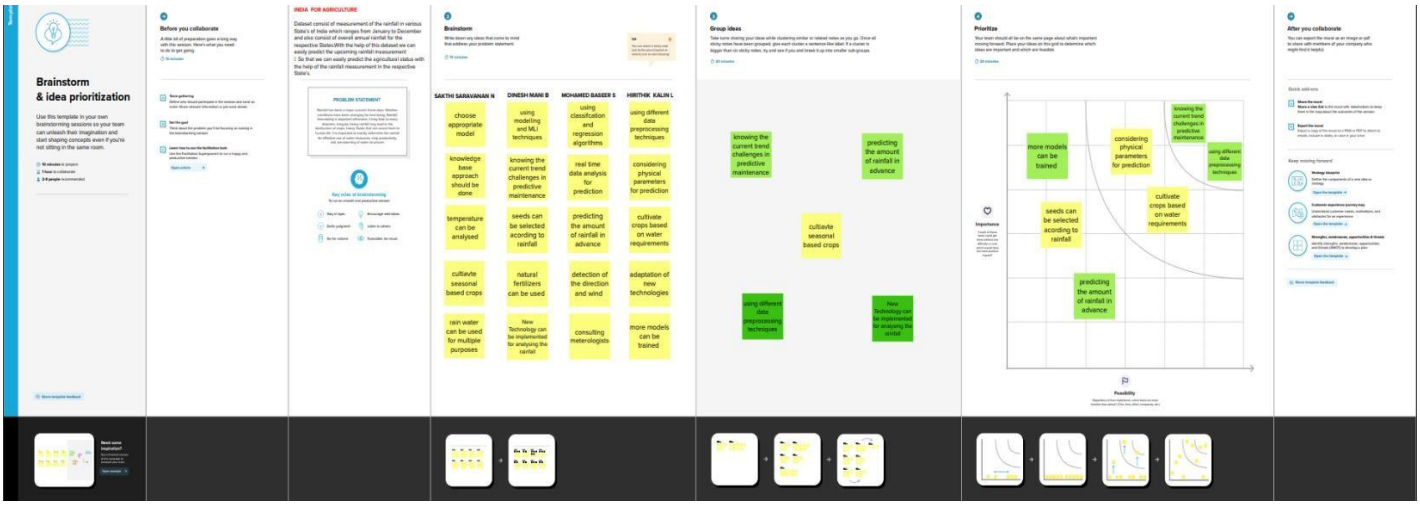
Rainfall has been a major concern these days. Irregular heavy rainfall may lead to the destruction of crops, heavy floods that can cause harm to human life. It is important to exactly determine the rainfall for effective use of water resources, crop productivity, and pre-planning of water structures..

3. IDEATION AND PROPOSED SOLUTION

3.1 Empathy Map Canvas:



3.2 Ideation and Brainstorming:

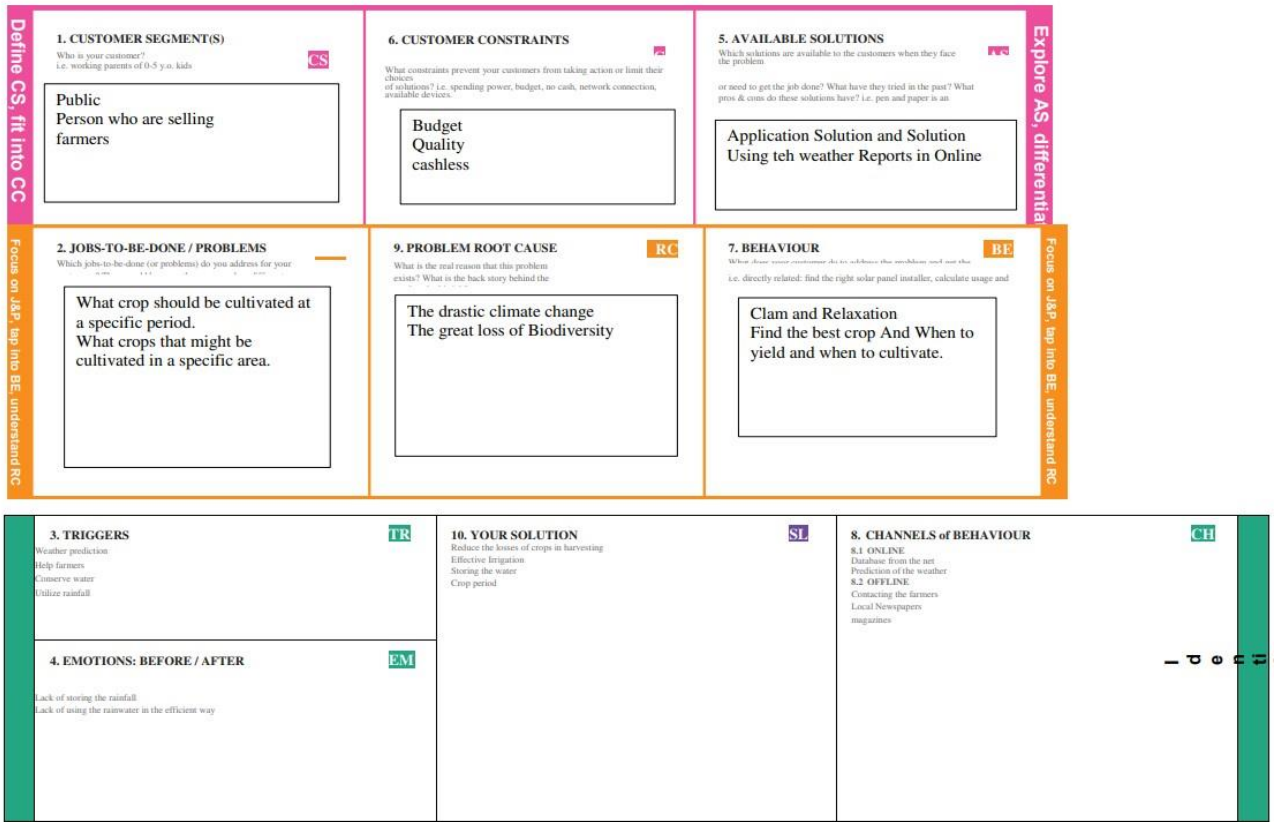


3.3 Proposed Solution:

The dataset consist of measurement of the rainfall in various State's of India which ranges from January to December and also consist of overall annual rainfall for the respective States With the help of this dataset we can easily predict the upcoming rainfall measurement ,So that we can easily predict the agricultural status with the help of rainfall measurement in the respective State's and machine learning algorithms.

3.4 Solution fit:

Project Title: Exploratory Analysis of RainFall Data in India for Agriculture **Project Design Phase-I - Solution Fit Template** **Team ID:** PNT2022TMID18172



4. REQUIREMENT ANALYSIS

4.1 Functional Requirements:

S.No	Component	Description	Technology
1.	User Interface	The user interacts with the application through a web UI and a chatbot	HTML, CSS, python, Flask
2.	Application Logic-1	Logic for registration Registration	Python
3.	Application Logic-2	Logic for login to theapplication	Python
4.	Application Logic-3	Integrating machine learning model and thewebpage	Flask
5.	Database	Numeric data	MySQL
6.	File Storage	To store files such as prediction report	Local Filesystem
7.	External API	Allows developers access to critical forecasts, alerts, andobservations, along with other weather data.	IBM Weather API

8.	Machine Learning Model	Predictive modeling is a statistical technique using machine learning and data mining to predict and forecast likely future outcomes with the aid of historical and existing data	Predictive modeling
9.	Infrastructure (Server)	Application Deployment on Local System Local Server Configuration: built-in flask web server	Flask web server

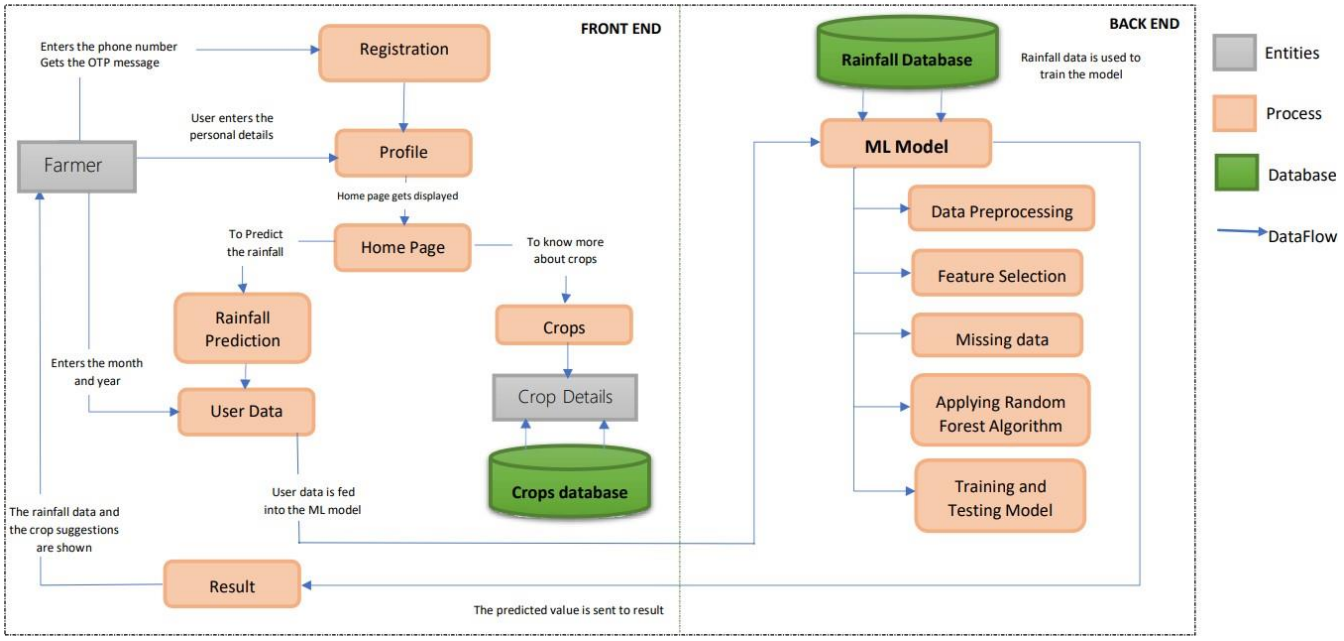
4.2 Non-Functional Requirements:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	Flask	Micro web framework written in Python
2.	Security Implementations	Basic HTTP authentication, Session based authentication, User Registration, Login Tracking	Flask Security
3.	Scalable Architecture	Size is everything, and Flask's status as a microframework means that you can use it to grow a tech project such as a web app incredibly quickly. Its simplicity of use and few dependencies enable it to run smoothly even as it scales up and up.	Flask

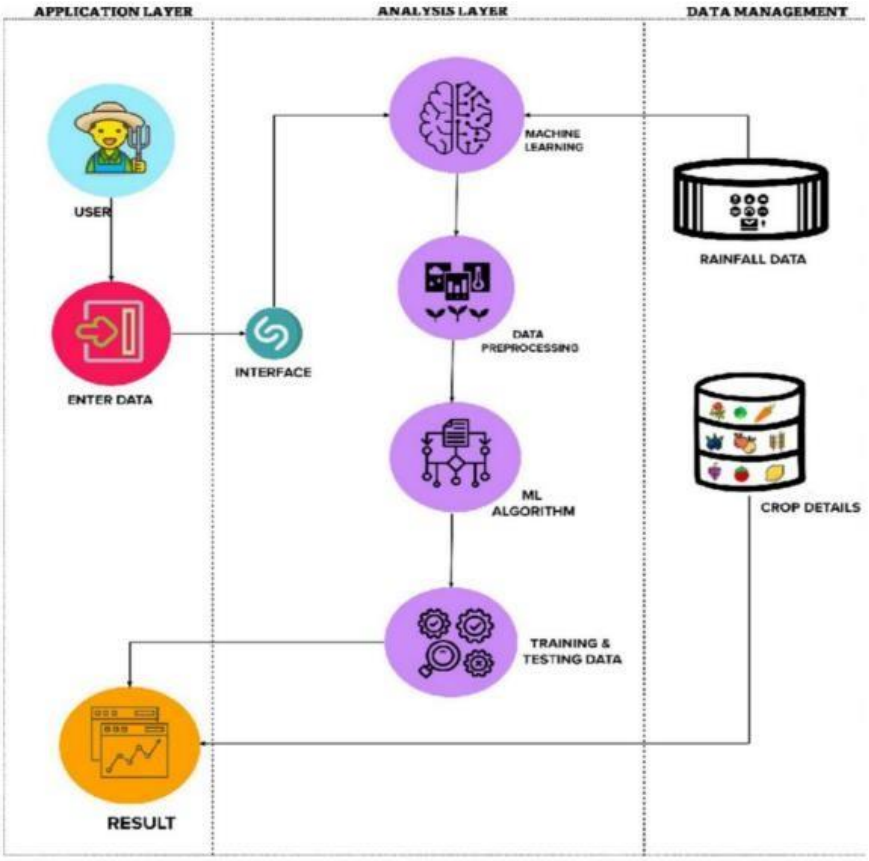
4.	**Availability**	**Higher compatibility with latest technologies and allows customization**	**Flask**
5.	**Performance**	• **Integrated support for unit testing.** • **RESTful request dispatching.** • **Uses Jinja templating.** • **Support for secure cookies (client side sessions)** • **100% WSGI 1.0 compliant.**	**Flask**

5. PROJECT DESIGN

5.1 Data Flow Diagram:



5.2 SOLUTION AND TECHNICAL ARCHITECTURE:



5.3 USER STORIES:

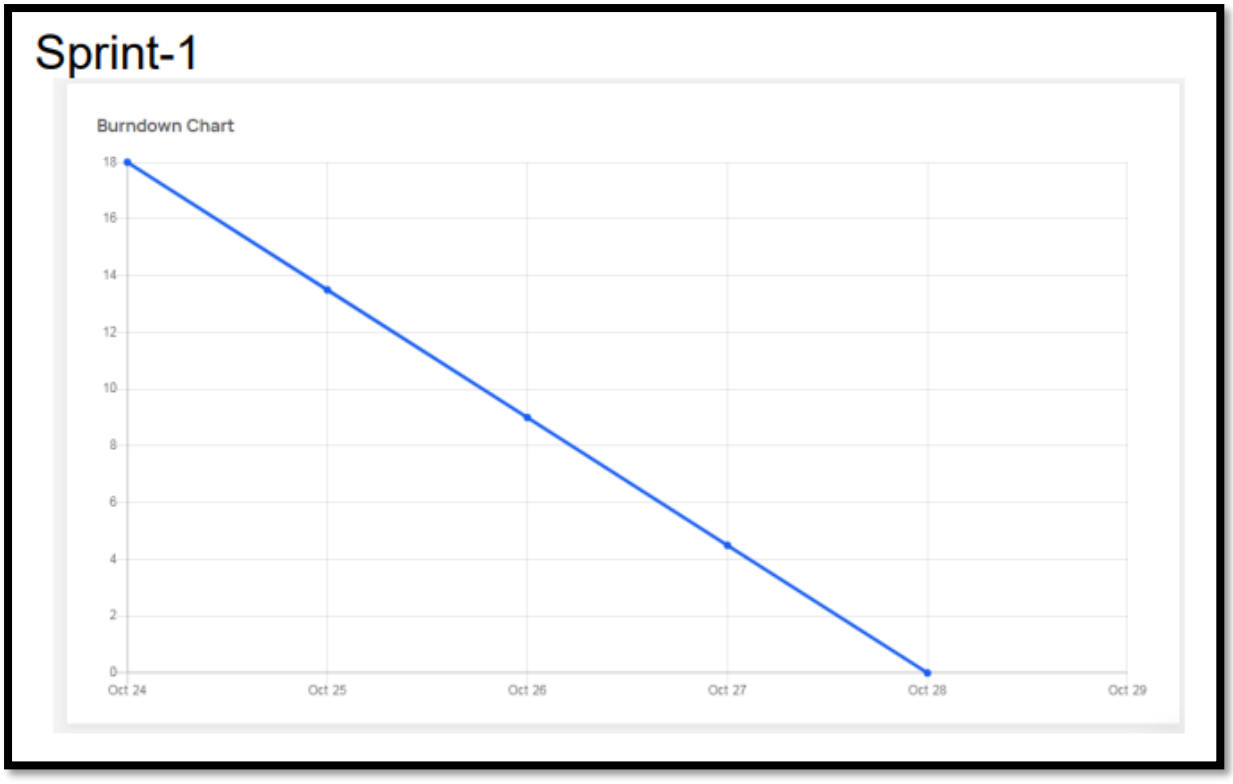
SPRINT NO	FUNCTION PERFORMED	TASK	STORY POINTS	PRIORITY	TEAM MEMBERS
SPRINT-1	Data Collection and processing	The needed data for processing and cleaning of the same has been done	20	High	L.Hirithik kalin, B.Dineshmani, S.Mohammed Baseer, N.Sakthi Saravanan
SPRINT-2	Model Building	Splitting of data into two sets and testing of the data is done	20	High	L.Hirithik kalin, B.Dineshmani, S.Mohammed Baseer, N.Sakthi Saravanan
SPRINT-3	Integrating with flask	Python has been integrated with flask framework	20	High	L.Hirithik kalin, B.Dineshmani, S.Mohammed Baseer, N.Sakthi Saravanan
SPRINT-4	Deployment of code	Deployment of code has been done	20	High	L.Hirithik kalin, B.Dineshmani, S.Mohammed Baseer, N.Sakthi Saravanan

6. PROJECT PLANNING & SCHEDULING:

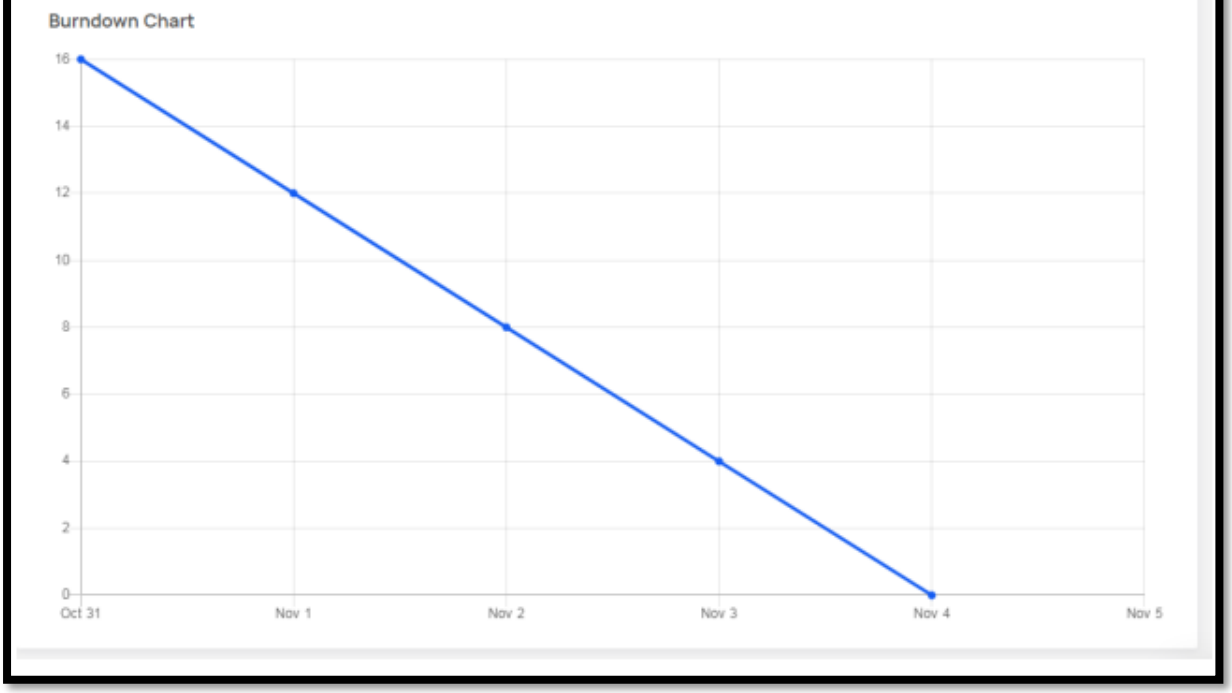
6.1 PLANNING AND ESTIMATION

Sprint	Total Story Points	Duration	Sprint StartDate	Sprint End Date (Planned)	Story Points Completed (Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

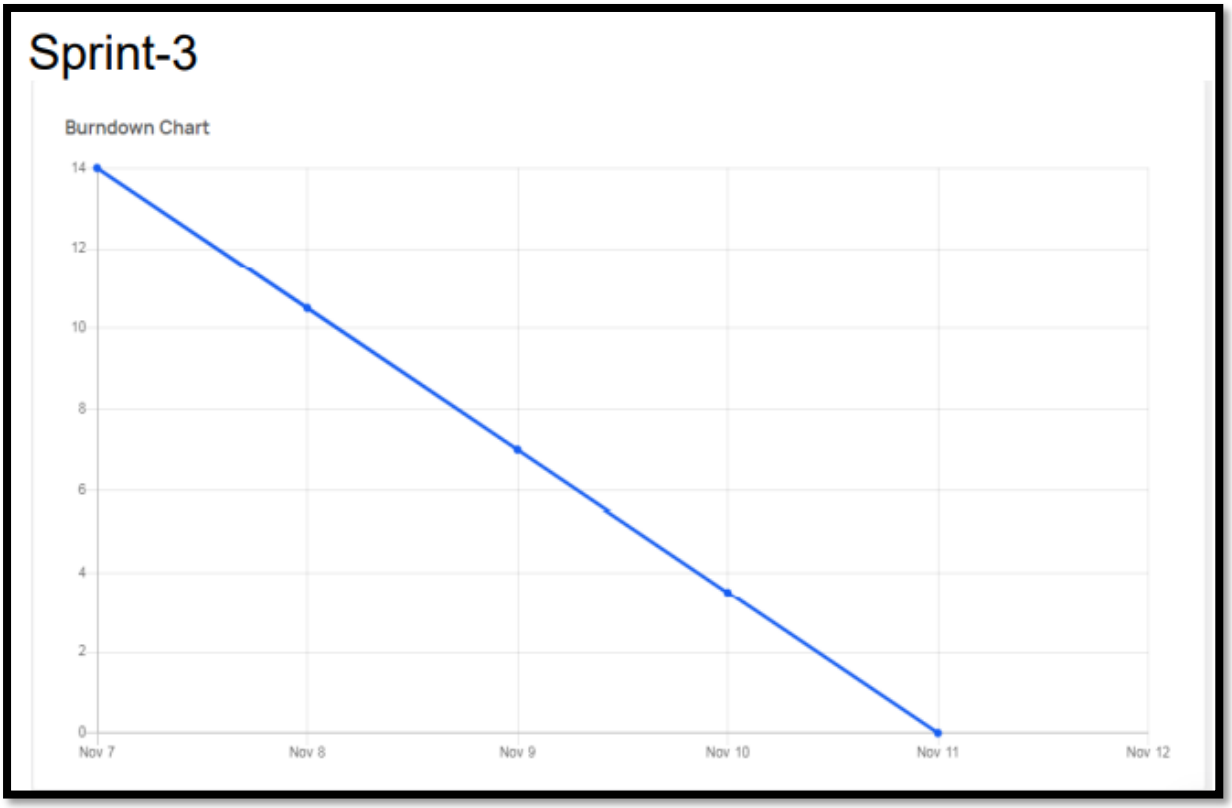
6.2 REPORTS FROM JIRA

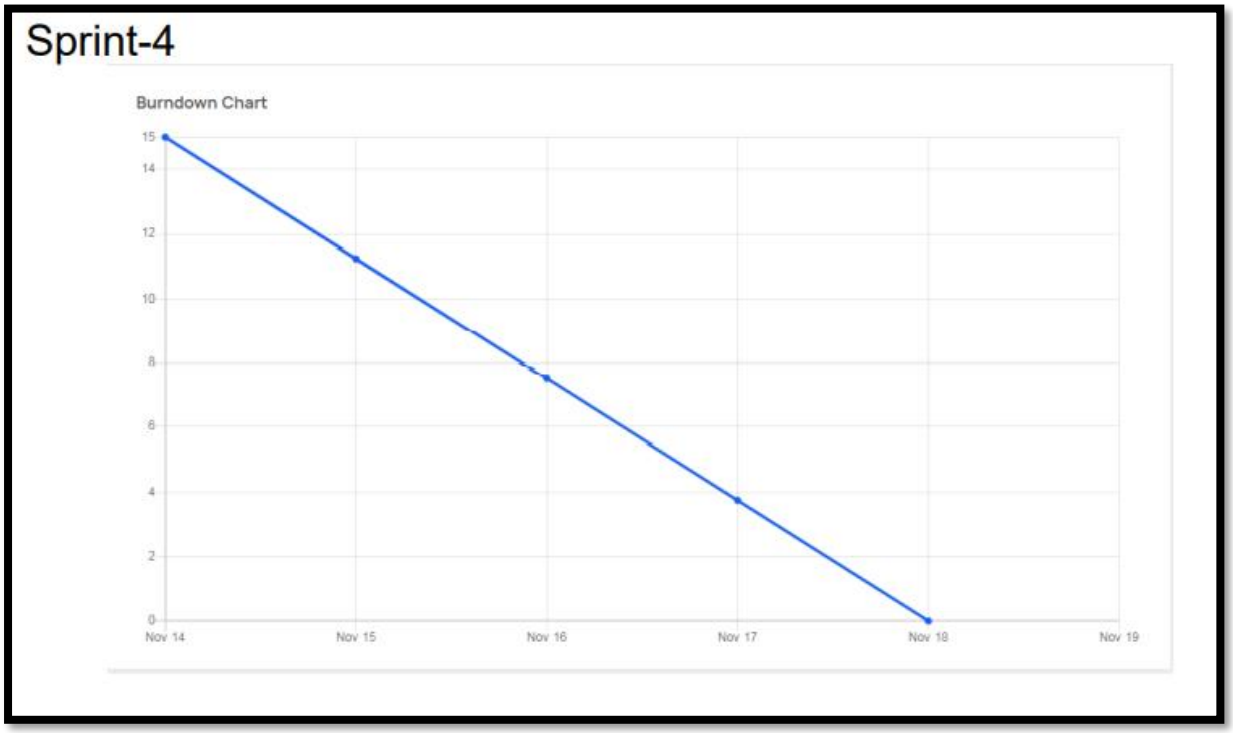


Sprint - 2



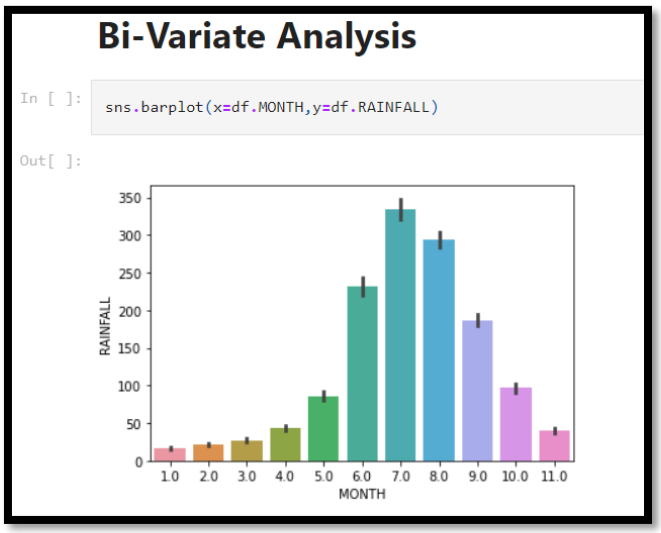
Sprint-3





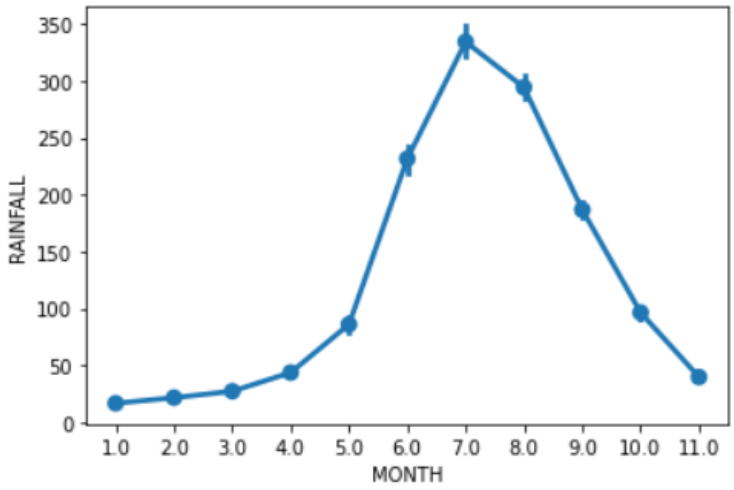
7. CODING & SOLUTIONING:

7.1 Feature 1



```
[ ]: sns.pointplot(x=df.MONTH,y=df.RAINFALL)
```

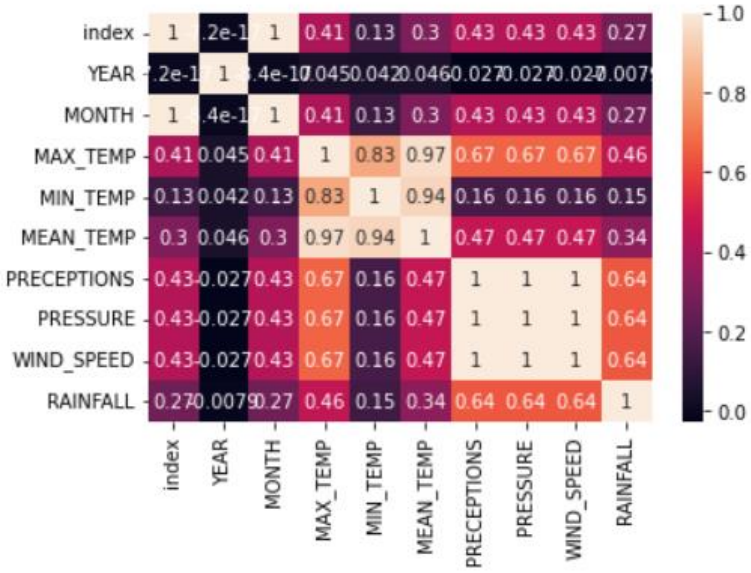
t[ ]:



7.2 Feature 2

```
In [ ]: sns.heatmap(s,annot=True)
```

Out[ ]:



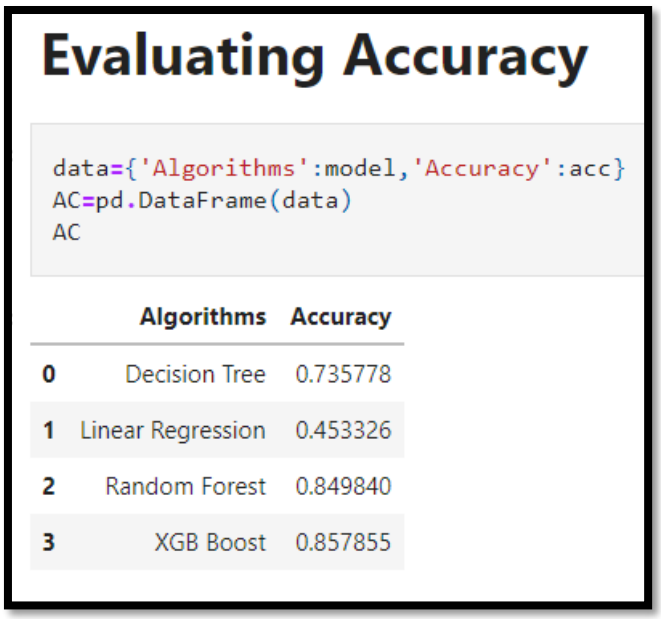
8. TESTING

8.1 Test Case Analysis:

This report shows the number of test cases that have passed, failed and untested.

Section	Total Cases	Not Tested	Fail	Pass
Home Page	2	0	0	2
Predict Page	4	0	0	4

8. RESULTS



10. ADVANTAGES & DISADVANTAGES

10.1 Advantage

1.Farmers can schedule the planting and harvesting of their crops.

2.When planning a vacation, people can choose a location and a time that will maximise the favourable weather.

3.Surfers are aware of when big waves are predicted.

4.If hurricanes or floods are predicted, certain areas can be evacuated.

5.Accurate weather predictions are extremely important for shipping and aircraft

10.2 Disadvantage

1.Weather is exceedingly challenging to predict accurately.
2.Monitoring so many different factors from so many different sources is expensive.
3.The cost of the computers required to do the requisite millions of calculations is high.
4. If the weather does not match the forecast, the weather forecasters are held responsible.

11. CONCLUSION

One of the most important things these days is the ability to predict the weather. To make risk management systems better and to automatically and scientifically predict the weather for the upcoming days. In order to help with weather prediction, numerous models have been developing.

Using six different machine learning (ML) algorithms-Cat Boost Classifier, Random For set Classifier, Logistic Regression, Gaussian NB, KNN, and XGB Classifier-we built a weather prediction web application from scratch in this study. The findings from all six models are listed in the results section, together with their accuracy, error rate, mean absolute error, root mean squared error, relative squared error, root relative squared error, and modelling time. The results demonstrate that, in comparison to all other classifiers utilized, the Cat Boost Classifier and the XGB Classifier have produced results with a high degree of accuracy. The Cat Boost Classifier surpasses

all other classifiers in tackling the problem at hand while taking the least amount of time to create the model.

12. FUTURE SCOPE

The WEATHER FORECASTING program will get new features in next releases, like: Live Location Tracking News on Current Disasters and deploy as an Android Application. Assistance in determining which crop will be most suitable given the weather.